EECS 225A Statistical Signal Processing	Lecture 7 - 02/11/2020

Lecture 7: Wiener Filter for Vectors and the Prediction Problem

Lecturer: Jiantao Jiao *Scribe: Shafeeq Ibraheem*

We first present solutions to the non-causal and causal Wiener filtering problems in the vector case instead of the WSS process case, and discuss the similarities of this setting to the WSS setting.

Then, we talk about the prediction problem in WSS processes and present a solution using spectral factorization. In this lecture, all the random variables are generally complex-valued.

1 Non-causal Wiener Filter in the Vector Case

To avoid the necessity of z-transform and convergence arguments, we consider the analogue of Wiener filtering in the vector case, where the signal we want to estimate is $X = (X_1, X_2, \ldots, X_N)^\top$ and the observation we have is $Y = (Y_1, Y_2, \ldots, Y_N)^\top$. Note that we have constrained that X and Y have equal length, since we want them to have shared time indices so that we can formulate a *causal* filter problem.

1.1 Non-causal Wiener filter (smoothing problem)

The natural question of non-causal Wiener filter would be to use the whole vector Y to estimate every entry of X. In other words, we would like to use estimator

$$\hat{X} = K_s Y \in \mathbb{C}^N$$

to estimate X. It follows from the orthogonality principle that for each $i, l \in [N]$,

$$\mathbb{E}[(X_i - \sum_{j=1}^{N} K_{s,ij} Y_j) Y_l^*] = 0,$$

which is equivalent to the matrix equation

$$R_{XY} = K R_Y,$$

where

$$\begin{aligned} R_{XY} &= \mathbb{E}[XY^*] \\ R_Y &= \mathbb{E}[YY^*], \end{aligned}$$

and $*$ denotes the conjugate transpose operation.

Clearly, if $R_Y > 0$, then K_s is given by

$$K_s = R_{XY} R_Y^{-1}.$$

Note that we *did not* assume X, Y are zero-mean.

1.2 Causal Wiener filter [KSH00, Chapter 4.1.2]

To estimate X_i, we would like to use

$$\hat{X}_i = \sum_{j=1}^{i} K_{f,ij} Y_j.$$

In other words, the estimate $\hat{X}_i$ is only allowed to *causally* depend on Y: it is a function of $Y_1, Y_2, \ldots, Y_i$ only. In other words, $K_{f,ij} = 0$ for any $j > i$.

It follows from the orthogonality principle that for any $i \in [N], 1 \leq l \leq i$,

$$\mathbb{E}[(X_i - \sum_{j=1}^{i} K_{f,ij} Y_j) Y_l^*] = 0,$$

which is equivalent to

$$R_{XY}(i, l) = \sum_{j=1}^{i} K_{f,ij} R_Y(j, l) \text{ for } 1 \leq l \leq i.$$

The $1 \leq l \leq i$ constraint tells us this relationship only holds for the lower triangular portion of these matrices. From this we derive the vector case Wiener-Hopf equation.

We introduce the operator $[]_L$: given a square matrix H, $[H]_L$ zeros out all of the entries of H in the strict upper triangular part. This is very similar to the $[]_+$ operation we used in the scalar causal Wiener filter.

We can succinctly write the Wiener-Hopf equation as

$$[R_{XY} - K_f R_Y]_L = 0$$

Analogously to the causal Wiener filtering solution using *spectral factorization* in WSS processes, here we use the *LDL decomposition* to solve it. The LDL decomposition is closely related to the Cholesky decomposition.

Theorem 1 (LDL decomposition)**.** *If matrix H is positive definite, then there exists a unique lower-diagonal-upper triangular factorization of H:*

$$H = LDL^*,$$

where L is lower triangular with unit diagonal entries, and D is diagonal with positive entries. Both L, D are invertible. Here $*$ is the conjugate transpose of a matrix, which is obtained from taking the transpose and then taking the complex conjugate of each entry.

We emphasize that the LDL decomposition solution to the causal filter problem here is not only conceptually important but also numerically efficient.

Now we continue to present the solution of causal filter problem. Writing the Wiener–Hopf equation without the $[]_L$ operator:

$$R_{XY} - K_f R_Y = U^+$$

Where U^+ is some strictly upper triangular matrix.

Assuming $R_Y > 0$, we have the LDL decomposition as

$$R_Y = LDL^*,$$

which implies

$$R_{XY} - K_f LDL^* = U^+.$$

Since L, D are invertible,

$$R_{XY} L^{-*} D^{-1} - K_f L = U^+ L^{-*} D^{-1}$$

Since K_f and L are both lower triangular matrices, $K_f L$ is also lower triangular. Since L is lower triangular, L^* is upper triangular–its inverse L^{-*} is upper triangular as well. Since D is diagonal and U^+ is strictly upper triangular, $U^+ L^{-*} D^{-1}$ is also strictly upper triangular. If we then apply the $[]_L$ operator to the above equation, we get:

$$[R_{XY} L^{-*} D^{-1}]_L - K_f L = 0$$

We have shown that

$$K_f = [R_{XY} L^{-*} D^{-1}]_L L^{-1}$$

How is it related to the non-causal Wiener filter? Indeed, in the non-causal case we have

$$\begin{aligned} K_s &= R_{XY} R_Y^{-1} \\ &= R_{XY} L^{-*} D^{-1} L^{-1} \\ &= K_f + [R_{XY} L^{-*} D^{-1}]_{\text{strict upper triangular part}} L^{-1} \end{aligned}$$

The interested readers must have observed that we have made a strong assumption here: $R_Y > 0$. Indeed, even for positive semidefinite matrices the LDL decomposition also exists (may not be unique), but it cannot be generally used to construct casual filter: in those cases L may not even have a *lower-triangular* pseudoinverse.

2 Prediction Problem

Consider a *zero-mean* WSS process X, and we would like to predict X_n using $X_{n-1}, X_{n-2}, ..., X_{-\infty}$. We can cast this as a causal Wiener filter with $Y_n = X_{n-1}$. The transfer function of the casual Wiener filter from Y to X is

$$H(\omega) = \frac{1}{S_Y^+(\omega)} \left[\frac{S_{XY}(\omega)}{S_Y^-(\omega)} \right]_+$$

Now it suffices to evaluate this formula in the special case of $Y_n = X_{n-1}$. We have

$$\begin{aligned} R_{XY}(k) &= \mathbb{E}(X_n Y_{n-k}^*) \\ &= \mathbb{E}(X_n X_{n-k-1}^*) \\ &= R_X(k+1) \\ R_Y(k) &= \mathbb{E}(Y_n Y_{n-k}^*) \\ &= \mathbb{E}(X_{n-1} X_{n-k-1}^*) \\ &= R_X(k) \end{aligned}$$

It implies that

$$\begin{aligned} S_{XY}(\omega) &= e^{j\omega} S_X(\omega) \\ S_Y(\omega) &= S_X(\omega) \end{aligned}$$

Plugging these into the formula:

$$\begin{aligned} H(\omega) &= \frac{1}{S_X^+(\omega)} \left[\frac{e^{j\omega} S_X(\omega)}{S_X^-(\omega)} \right]_+ \\ &= \frac{1}{S_X^+(\omega)} [e^{j\omega} S_X^+(\omega)]_+ \end{aligned}$$

We can simplify this further by looking at $S_X^+(\omega)$. We know this is a causal signal, and can be represented as:

$$S_X^+(\omega) = p_0 + \sum_{k\geq 1} p_k e^{-j\omega k}, \tag{1}$$

where $p_0 = \sqrt{r_e} > 0$. Hence,

$$[e^{j\omega} S_X^+(\omega)]_+ = (S_X^+(\omega) - p_0)e^{j\omega}$$

Plugging this back into the formula for $H(\omega)$, we obtain

$$H(\omega) = (1 - \frac{p_0}{S_X^+(\omega)})e^{j\omega}$$

We now have the transfer function for optimal causal filter from Y to X. However, the input to the filter is X instead of Y, so the output should be delayed by one unit, resulting in the optimal *prediction* filter:

$$H(\omega) = (1 - \frac{p_0}{S_X^+(\omega)}).$$

It can be shown that if we would like to predict X_n using all the information up to $X_{n-\tau}$, the optimal prediction filter is given by

$$H_\tau(\omega) = \frac{\sum_{k=\tau}^{\infty} p_k e^{-j\omega k}}{S_X^+(\omega)},$$

where $S_X^+(\omega)$ is given by (1).

We finish our discussions on the prediction filter by computing the spectrum of the error process $E_n = X_n - \hat{X}_n$. This can be understood as passing X through a filter $1 - H(\omega)$. It implies that

$$\begin{aligned} S_E(\omega) &= |1 - H(\omega)|^2 S_X(\omega) \\ &= |1 - (1 - \frac{p_0}{S_X^+(\omega)})|^2 S_X(\omega) \\ &= \frac{p_0^2}{|S_X^+(\omega)|^2} S_X(\omega) \\ &= p_0^2. \end{aligned}$$

We see $S_E(\omega)$ is a constant, which means that the process E_n is a white noise process with variance p_0^2. That is why in the literature the process E_n is called *innovation* process: it captures the *innovation* part of X_n given all of its history up to times $n-1$, and is a white noise process.

3 Generalizations of the prediction problem

Consider the following generalizations of the prediction problem. Say we have

$$Y_i = S_i + V_i, \tag{2}$$

where $\langle V_i, V_j \rangle = r\delta_{ij}, \langle V_i, S_j \rangle = 0$. In other words, $S_V(z) = r, S_{VS}(z) = 0$. Furthermore, we have $S_{SY}(z) = S_S(z), S_Y(z) = S_S(z) + S_V(z) = S_S(z) + r$. Our goal is to estimate S_i using all the information of $\{Y_j\}_{j\leq i}$. It is a causal Wiener filter problem, and the solution is given by

$$K_f(z) = \left[\frac{S_{SY}(z)}{L^*(z^{-*})}\right]_+ \frac{1}{r_e L(z)}, \tag{3}$$

where $S_Y(z) = L(z)r_e L^*(z^{-*})$ is the canonical spectral factorization of Y. We can simplify it by

$$K_f(z) = \left[\frac{S_{SY}(z)}{L^*(z^{-*})}\right]_+ \frac{1}{r_e L(z)} \tag{4}$$

$$= \left[\frac{S_Y(z) - r}{L^*(z^{-*})}\right]_+ \frac{1}{r_e L(z)} \tag{5}$$

$$= \left[L(z)r_e - \frac{r}{L^*(z^{-*})}\right]_+ \frac{1}{r_e L(z)} \tag{6}$$

$$= \left([L(z)]_+ - \left[\frac{1}{L^*(z^{-*})}\right]_+ \frac{r}{r_e}\right) \frac{1}{L(z)} \tag{7}$$

Since $L(z)$ is causal, we have $[L(z)]_+ = L(z)$. Since $\frac{1}{L^*(z^{-*})}$ is anticausal and $L^*(z^{-*})|_{z=0} = 1$, we have $\left[\frac{1}{L^*(z^{-*})}\right]_+ = 1$. Then we have

$$K_f(z) = 1 - \frac{r}{r_e} L^{-1}(z). \tag{8}$$

One might wonder about the filter that yields the optimal estimate of S_i given $\{Y_j\}_{j \le i-1}$. Noting that the optimal linear estimator of S_i given this history is the same as the optimal linear estimator of Y_i given this history, so this problem is equivalent to the canonical prediction problem we just studied, which has filter

$$K_p(z) = 1 - L^{-1}(z). \tag{9}$$

Note that both filters take the Y process as input and outputs

$$(h * Y)_n = \sum_{j=-\infty}^{\infty} h(j) Y_{n-j}. \tag{10}$$

References

[KSH00] Thomas Kailath, Ali H Sayed, and Babak Hassibi. *Linear estimation*. Number BOOK. Prentice Hall, 2000.